Combined Parameter and State Estimation by Two-stage Filtering

Naratip Santitissadeekorn and Chris Jones

Department of Mathematics
University of North Carolina-Chapel Hill

Nonlinear filtering and Data Assimilation Workshop,
Bangalore, 11 January 2014

ENKF and PF

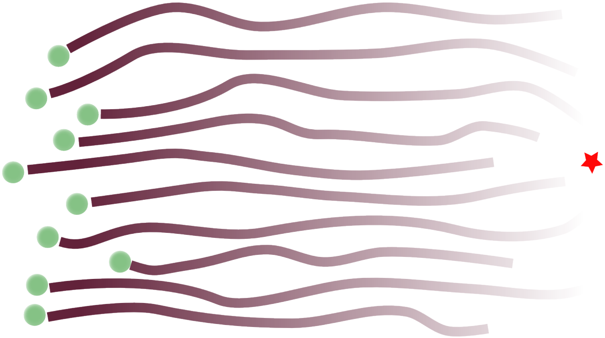
- **ENKF**: Approximate mean and covariance of $P(x_t|y_{1:t})$
  - Using ensemble to approximate statistics of the forecasts/background
  - Using Kalman update eqs. to approximate posterior statistics
  - Inflation: increasing the background error covariance
  - Localization: suppressing spurious correlation from distant states
- **PF**: Empirical approximation of $P(x_t|y_{1:t}) \approx \{w_t^{(i)}, x_t^{(i)}\}$
  - No update for $x_t^{(i)}$ only $w_t^{(i)}$

$$w_t^{(i)} \propto \frac{P(y_t|x_t^{(i)})P(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{1:t-1}^{(i)}, y_{1:t}))} w_{t-1}^{(i)}$$

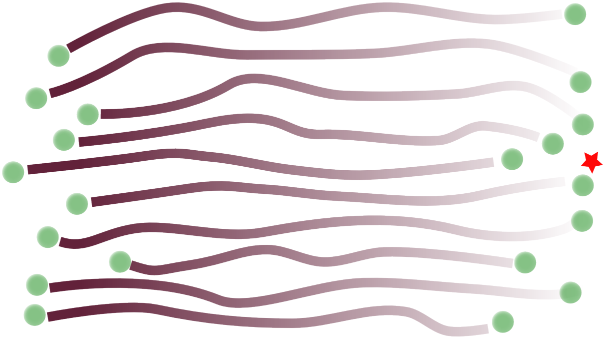
  - Resampling: focus computation on particles with high weight
  - Jittering or Merging: deal with the 'impoverishment' problem

Standard PF (in a nutshell)

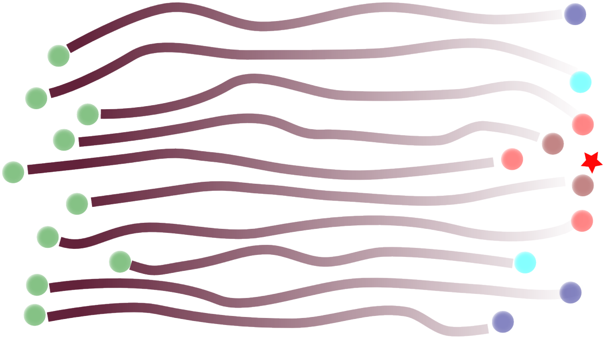
Standard PF (in a nutshell)



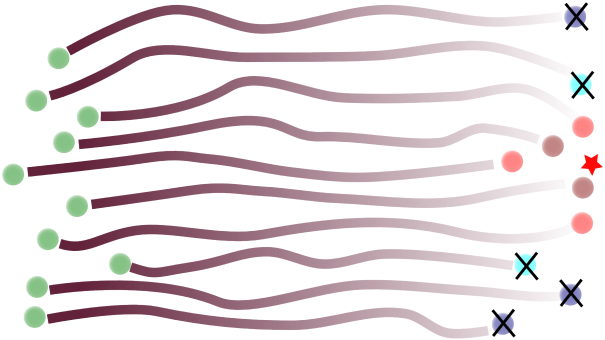
Standard PF (in a nutshell)



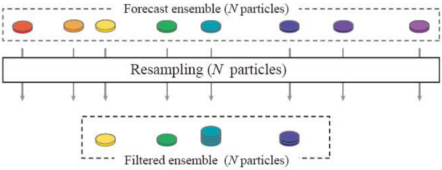
Standard PF (in a nutshell)



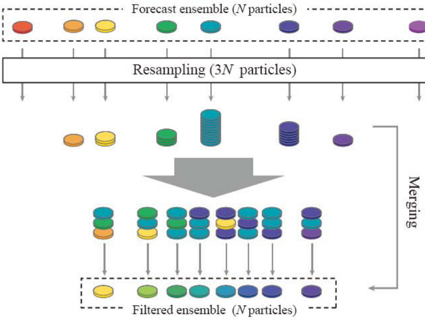
Standard PF (in a nutshell)



Merging [Nakano et al., Nonlin. Process. Geophys. 2007]



Particle Filter (PF): A filtered ensemble is obtained by simple resampling.



Merging Particle Filter (MPF):

Experimental Setup: Lorenz-96

- **True model**: 40 variables system of Lorenz-96

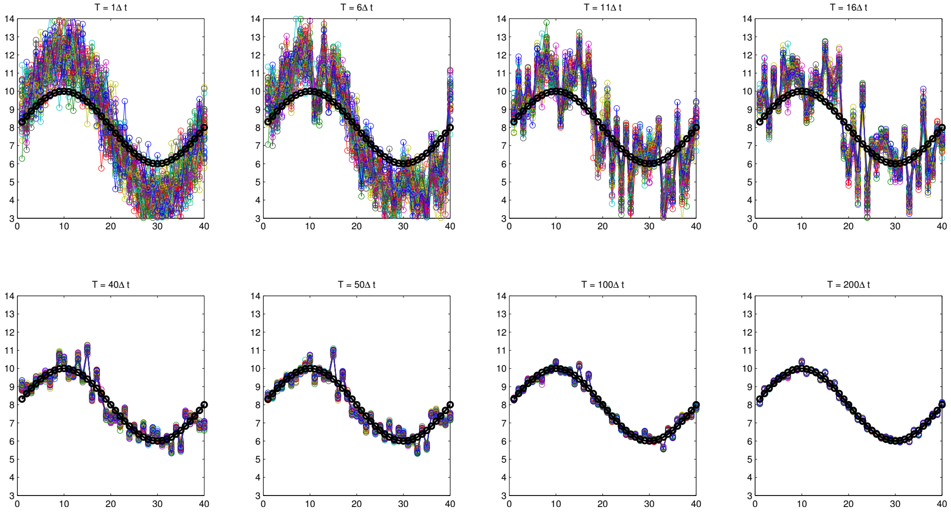
$$\dot{x}_i = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F_i,$$

- $i = 1, \ldots, 40$ with cyclic boundary conditions
- Forcing parameter F_i vary with i
- Numerically solved with $\Delta t = 0.05$
- **True Parameters**:$F_i = 8 + 2\sin(\frac{2\pi}{40} i)$
- 30,000 spin up run and the run from 30,001-36,000 will be used in this experiment

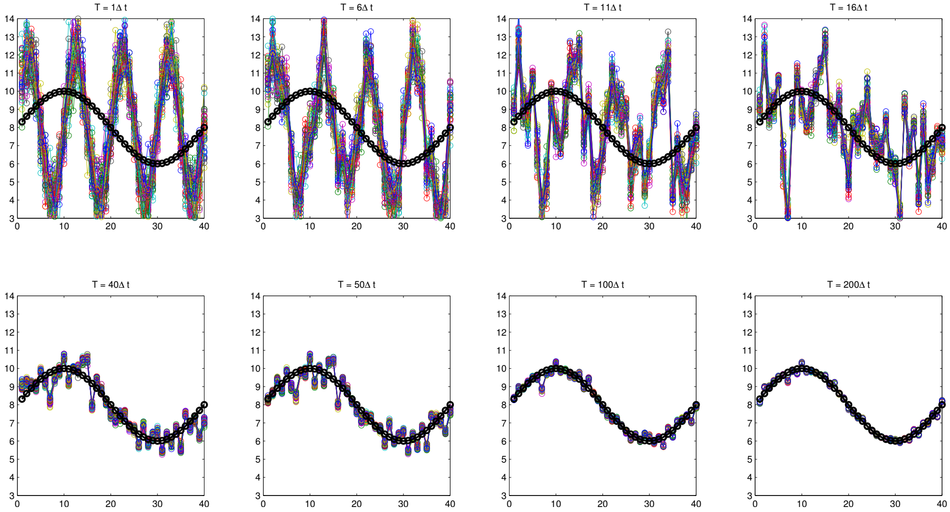
Experimental Setup: ENKF

- **State Augmentation**: $[x_i, F_i]$
- **Observation**: $y = [x_1, x_3, \ldots, x_{39}] + \mathcal{N}(0, 0.1)$
- **Localization**: Covariance filtering, Gaspari-Cohn
- Assume a perfect (forecast) model scenario (no stochastic term)
- The interval between observations,δ, controls the degree of nonlinearity
- For $\delta = \Delta t$, approximately linear system map
- For $\delta \geq 5\Delta t$, the forward map becomes apparently nonlinear [Bengtsson et al. 2003]

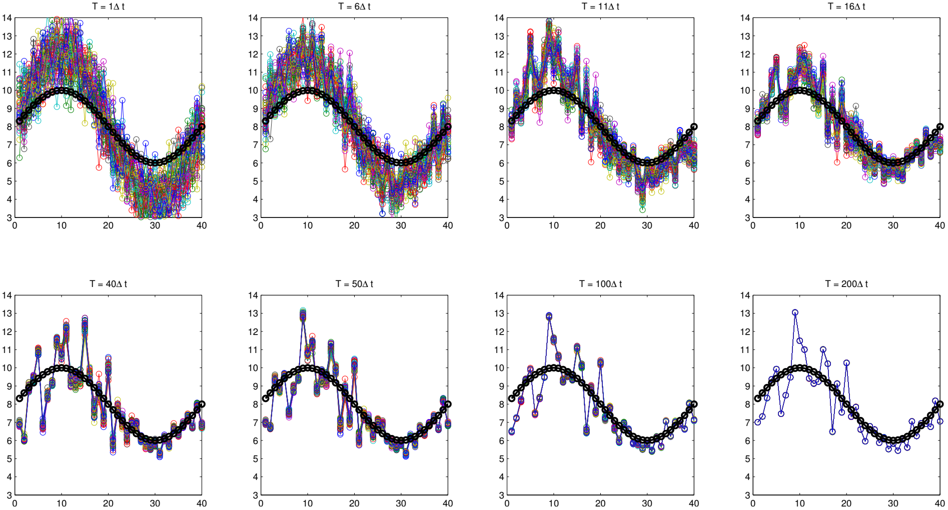
Results: ENKF with $N = 50, \delta = \Delta t$



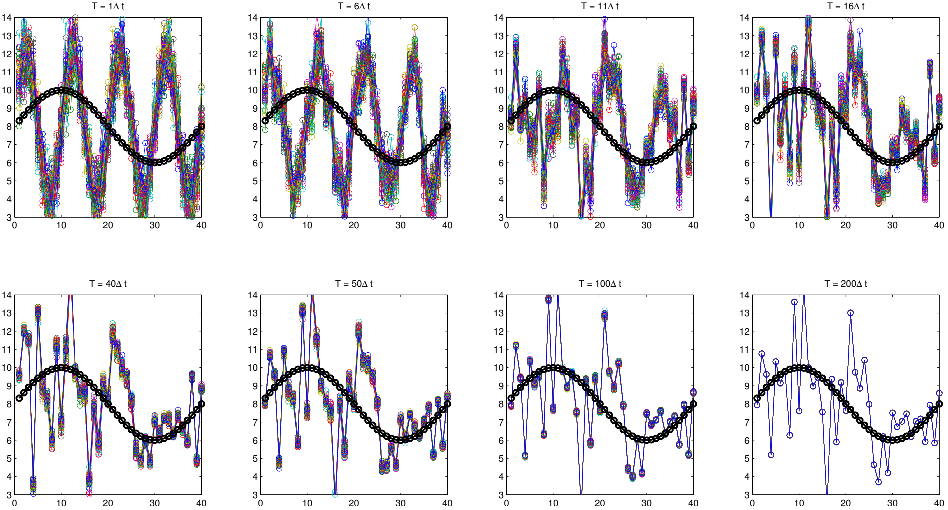
Results: ENKF with $N = 50, \delta = \Delta t$



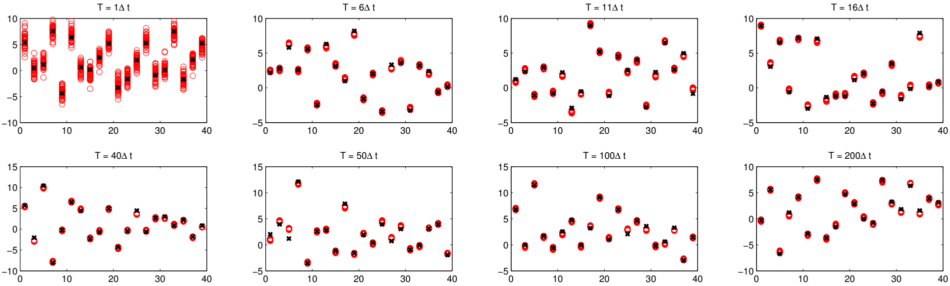
Results: ENKF with $N = 50, \delta = 10\Delta t$



Results: ENKF with $N = 50, \delta = 10\Delta t$



Results: ENKF with $N = 50, \delta = 10\Delta t$



Two-stage Filter: PF+ENKF

- **Motivation**: Based on the Lorenz-96 experiment
  - Inferring about forcing fails for a long observation interval
  - But ENKF can still keep track of the true state variable
- **Idea**: Run PF for parameter (assuming known states), Run ENKF for states (assuming known parameters)
- **Parameter filter with PF**: use the state estimate $\hat{x}_{t-1}$

$$\theta_t = \alpha\theta_{t-1} + (1-\alpha)\overline{\theta}_{t-1} + \eta_t \qquad 0 < \alpha < 1$$
$$y_t = h(\phi_{t-1}^t(\hat{x}_{t-1};\theta_t);\theta_t) + \zeta_t$$

- **State filter with ENKF**: use the parameter estimate $\hat{\theta}_t$

$$x_t = \phi_{t-1}^t(x_{t-1};\hat{\theta}_t) + \nu_t$$
$$y_t = h(x_t;\hat{\theta}_t) + \eta_t$$

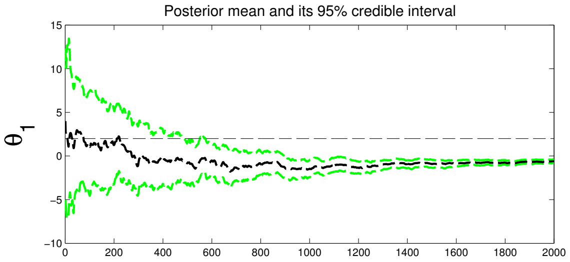
Shrinkage of Kernel

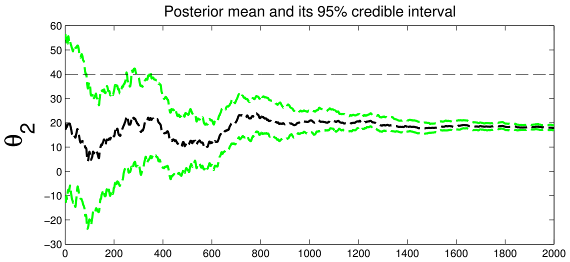
- **Why not persistence model?**: $\theta_t = \theta_{t-1}$
  - "Sample attrition" issue in re-weighting
  - (Parameter) particles always stay with the same sets of initial guesses
- **Random Walk**: $\theta_t = \theta_{t-1} + \eta_t$
  - Lead to over-dispersion; posteriors are far too diffuse relative to the theoretical posteriors for the actual fixed parameters
  - This over-dispersion is quantified in the framework of kernel smoothing [Liu&West 99]
- **Shrinkage of Kernel**: $\theta_t = \alpha\theta_{t-1} + (1-\alpha)\overline{\theta}_{t-1} + \eta_t$
  - Push samples θ_t toward their mean before adding a small degree of noise
  - $0.9 \leq \alpha < 1$

Two-stage Filter: PF+ENKF

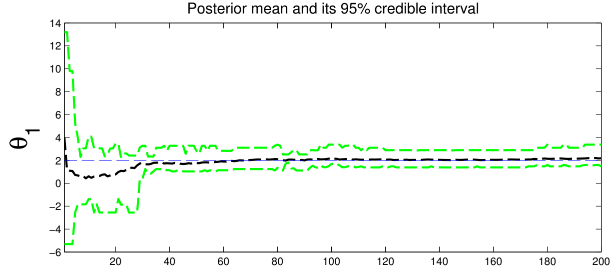
- Does it works for the Lorenz-96 experiments? No! Why?
  - PF suffers from the curse of dimensionality
  - Parameter space is 40-dimensional in this Lorenz-96 example!
- **Parameterized forcing parameter**: Assume $F = f_0 + \theta_1 \sin(\frac{2\pi}{\theta_2} i)$
- Assume $f_0 = 8$ is known and try to estimate $\theta = [\theta_1, \theta_2]$
- True parameter: $\theta^* = [2, 40]$
- Initial samples $\theta_0 \sim N([4, 20], \Lambda([10, 100]))$
- Can θ be linearly regressed (ENKF) from the observation of x_i?
- Will the two-stage filtering become more effective in this case?

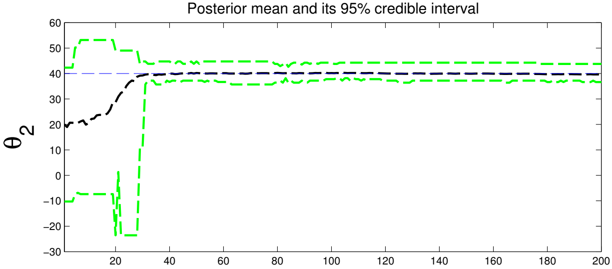
Results: ENKF with $N = 250, \delta = \Delta t$



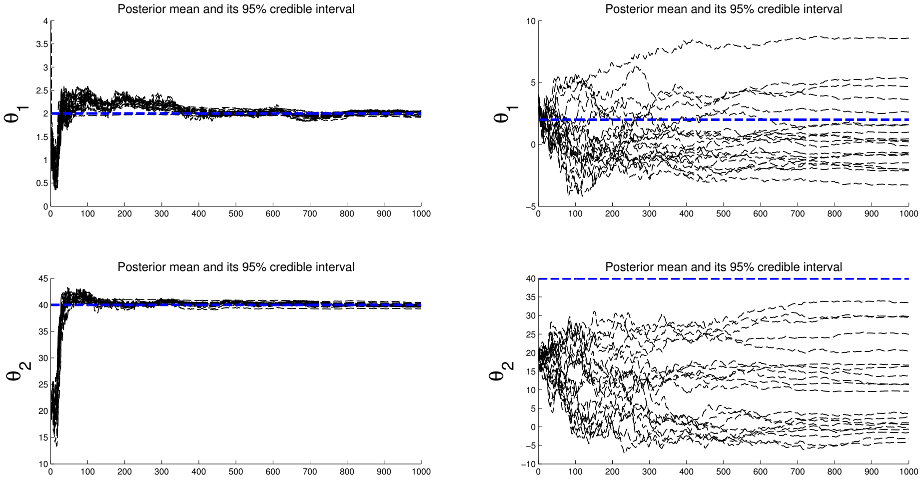


Results: Two-stage filtering with $N = 200 : 50, \delta = \Delta t$

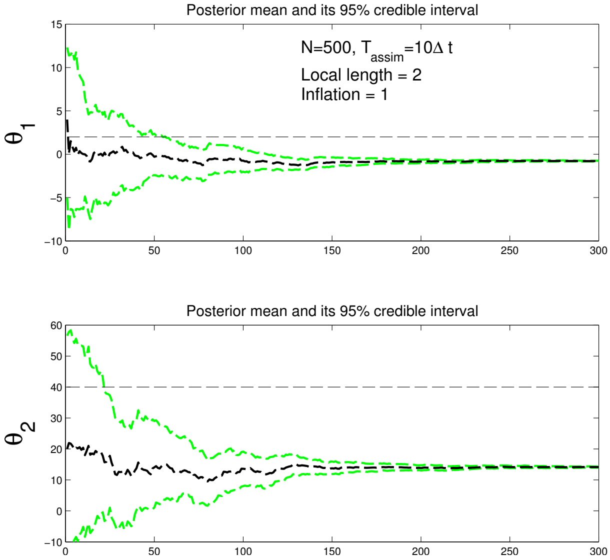




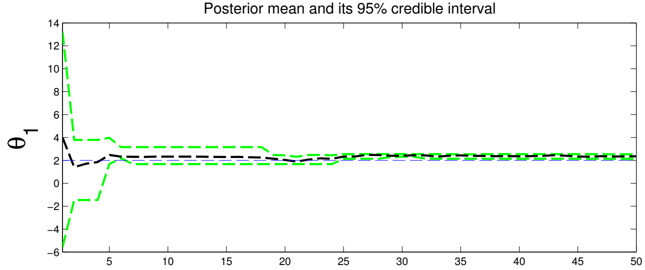
Results: 50 experiments, $N = 200 : 50, \delta = \Delta t$

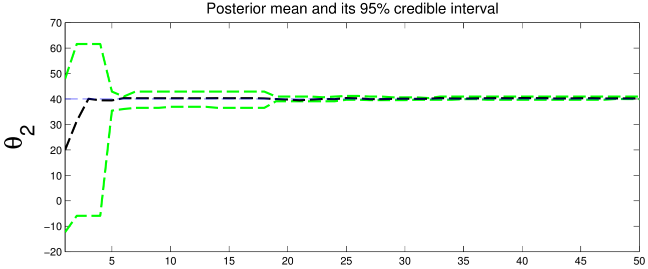


Results: ENKF with $N = 500, \delta = 10\Delta t$

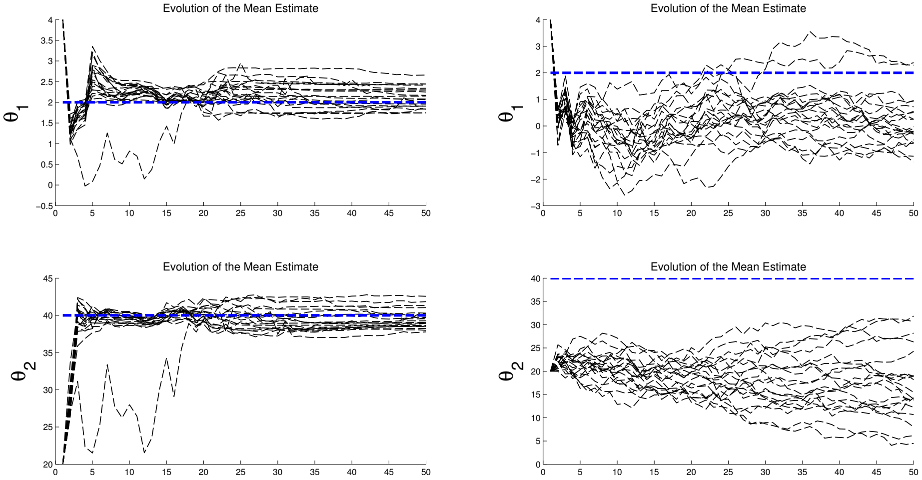


Results: Two-stage filtering with $N = 450 : 50, \delta = 10\Delta t$





Results: 50 experiments, $N = 200 : 50, \delta = \Delta t$



Summary: Two-stage Filter (PF-ENKF)

- **Advantage**
  - more effective in the cases where parameters cannot be linearly regressed from state estimates (i.e. rely only on the covariance between model states and parameters)
  - Minimal code modification is needed
- **Limitation**
  - Parameter space must be “small enough”
- **Future work**
  - More experiments: Fast(atmosphere)-Slow(ocean) system (e.g. the model in Zhang et al. 2012)
  - effective for temporally varying parameters?
  - parameter estimation in Lagrangian DA
  - Enough “information” in Lagrangian observations to allow parameter inference